



# ReReco @ CCIN2P3



## 2 basic building blocks:

**1. SAM** – Data Management & Transfer  
FNAL <--> CCIN2P3

## 2. Local Job & Metadata Management

- originally developed for MC-production (Patrice)
- based on Oracle DB & Python scripts
- submission
- monitoring
- book-keeping



# SAM – 1



## Data Transfer FNAL → CCIN2P3

### – Sam project

→ used bbftp for storing directly into HPSS

### – Locally running process check files

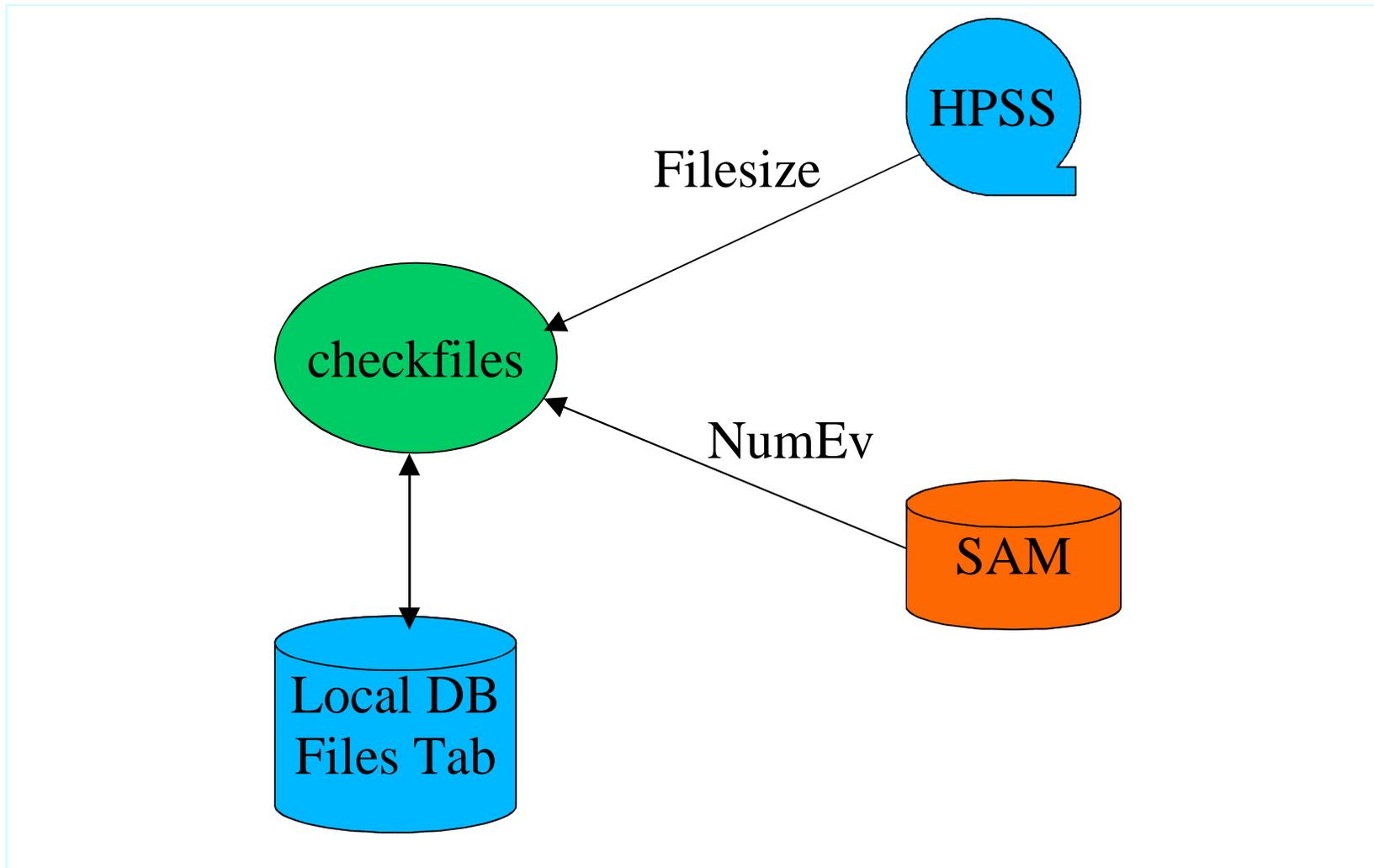
HPSS(filesize) + SAM(numev)

### → Local DB – Files Table

– Jobname, Type, Sam, Status, Location\_Status,  
Filename, Location, Numev, Filesize, Parent



# Local Process – checkfiles





## 1. Assign to each file a job

--> **Jobs Table (DB):**

**RequestID** – for each dataset unique,

**Jobname** – unique,

**Jobtype** (d0reco, bbftp),

**Status** (pending, running, done, error)

**RefJobname** ...

--> **Jobs Description Table :**

**RequestID, Jobtype, Macro, HPSS–location, BQSarg ...**



## 2. Running process – submitting, checking jobs

--> updating Jobs Table

- **check:** -- matching BQS-status <--> local DB

-- metadata

-- return code from executable

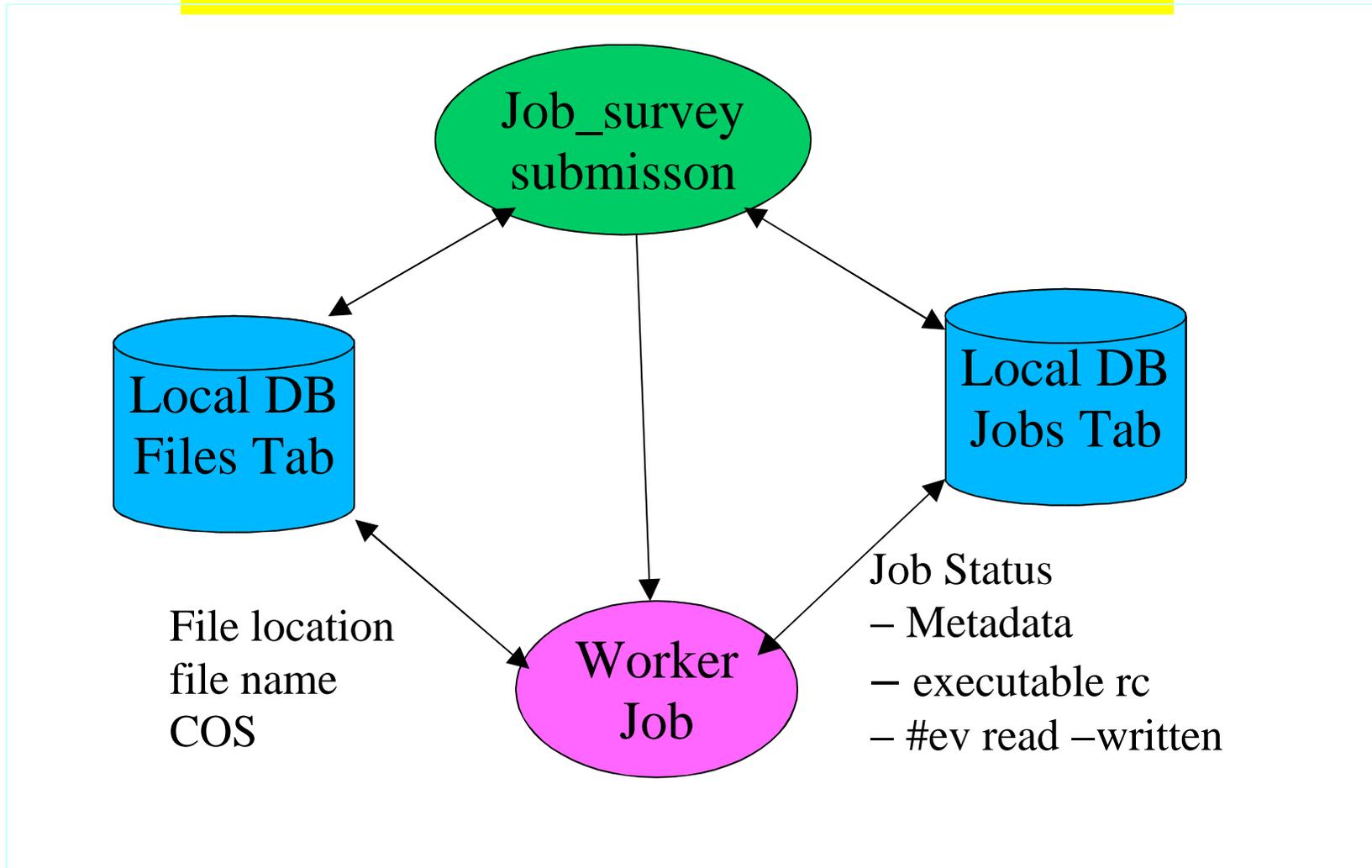
-- #events read <--> # events written

**any error** --> set error flag in Jobs Table

--> clean, resubmit job

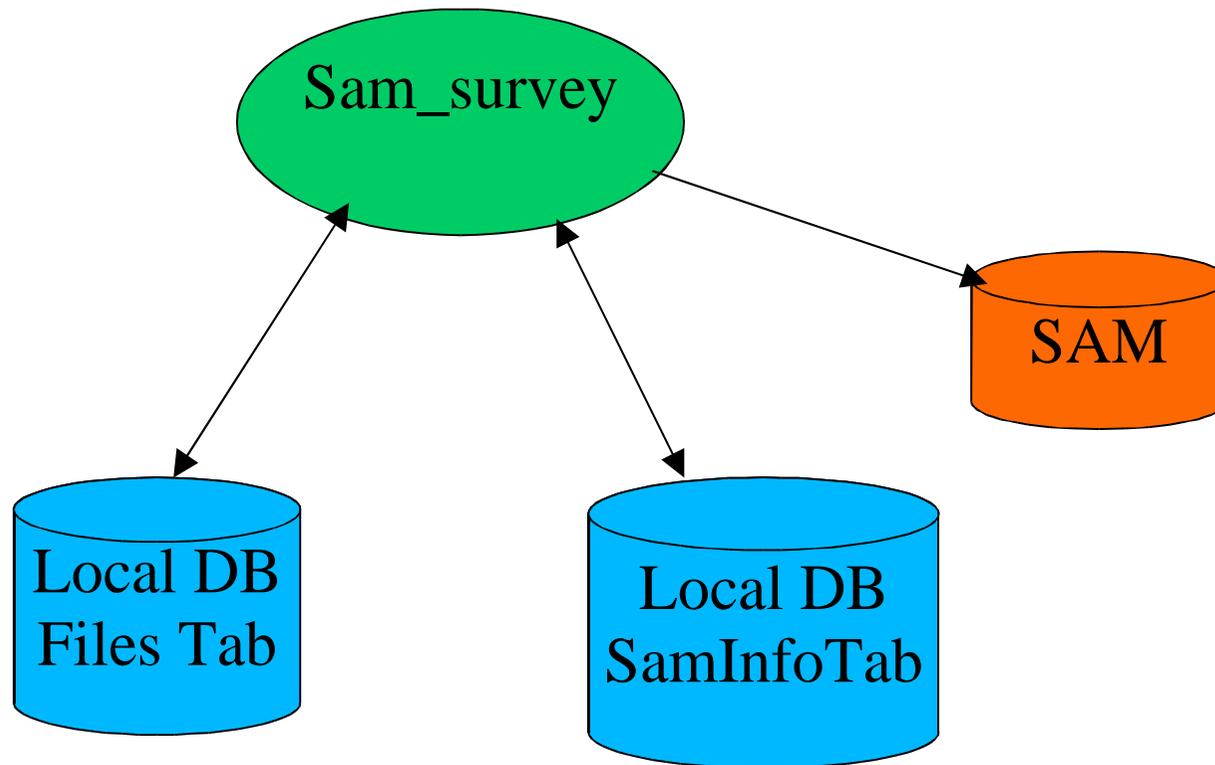


# Local Process – Job submission





## 1. SAM declaration, location (after each file done) – DSTs only





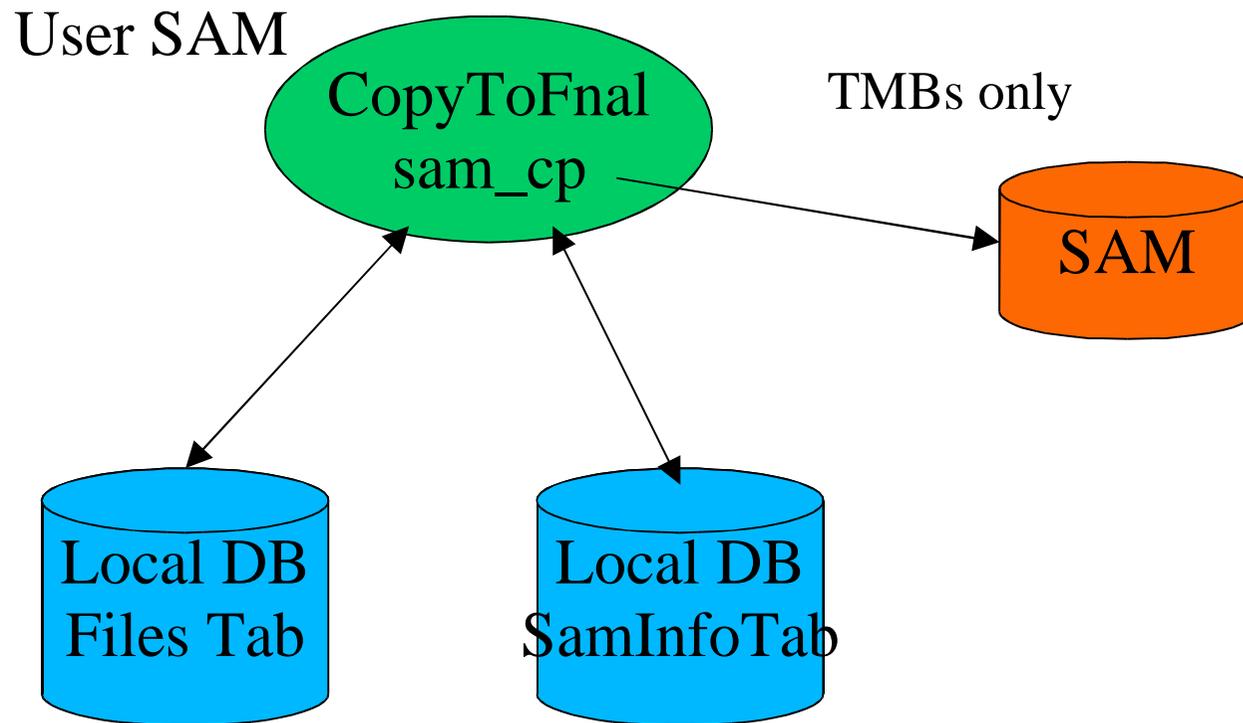
## 2. Transfer CCIN2P3 --> FNAL

- check Local DB
- if all files of a given project reprocessed
- > TMBs copied to FNAL , deleted @ IN2P3
  - bbftp faster then gridftp
- DSTs not transferred, stored @ CCIN2P3

----> web monitoring based on the local DB information



# Local Process – CopyToFnal





# Summary



- **decoupled reprocessing phases** (... and problems)
- **status of each phase is written into local DB**
- **error flags** --> **simple cleaning & jobs resubmission**
- **archives of logfiles & of metadata are stored in HPSS**
  
- > **monitoring of the actual status**
- > **statistics** (file size, cpu-time, processor used, # events ...)
- > **easy identification of problems**
- > **system tolerant to errors in different phases**
- > **very efficient reprocessing with relatively low CPU power**
  
- ... **at the end of the day we have our own “microSAM”**

